



# The predictive accuracy of credit ratings: Measurement and statistical inference

Walter Orth\*

Department of Statistics and Econometrics, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany

## ARTICLE INFO

### Keywords:

Ratings  
Evaluating forecasts  
Accuracy Ratio  
Harrell's C  
Overlapping lifetimes

## ABSTRACT

Credit ratings are ordinal predictions of the default risk of an obligor. The most commonly used measure for evaluating their predictive accuracy is the Accuracy Ratio, or equivalently, the area under the ROC curve. The disadvantages of these measures are that they treat default as a binary variable, thus neglecting the timing of default events, and they fail to use all of the information available from censored observations. We present an alternative measure which is related to the Accuracy Ratio but does not suffer from these drawbacks. As a second contribution, we study statistical inference for the Accuracy Ratio and the proposed measure in the case of multiple cohorts of obligors with overlapping lifetimes. We derive methods which use more sample information and lead to tests which are more powerful than alternatives which filter just the independent part of the dataset. All procedures are illustrated in the empirical section using a dataset of S&P Credit Ratings. © 2011 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

## 1. Introduction

Ratings are ordinal predictions of the default risk of an obligor. As with any prediction problem, the evaluation of the predictive accuracy is an essential factor. The measure which is most commonly used by rating agencies, regulators and researchers is the Accuracy Ratio, which is the summary statistic of the so-called Cumulative Accuracy Profile.<sup>1</sup> For the calculation of the Accuracy Ratio, it is necessary to choose a fixed time horizon and classify all of the obligors into two groups, those who defaulted within the chosen time span and those who did not. However, reducing the data to this kind of classification leads to a loss of information. First, the timing of the defaults is neglected. Second, certain right-censored observations

have to be omitted. The latter concerns those obligors who are observed – without default – for only a fraction of the chosen prediction horizon. The proportion of these kind of right-censored observations, and thus the loss of information, grows with the prediction horizon. In this paper we show how we can extend the methodology of the Accuracy Ratio to include all of the information contained in rating datasets. We do so by introducing to the credit risk literature 'Harrell's C', a so-called concordance index which has been proposed in the biostatistical literature for the purpose of evaluating the predictive accuracy (Harrell, Lee, & Mark, 1996). It should be clarified that Harrell's C is not an improved estimator of the population Accuracy Ratio. Instead, we argue that the population Harrell's C is a better measure of the predictive accuracy. Furthermore, we also propose a modification of Harrell's C which takes the prediction horizon into account and is expected to be more suitable for credit risk applications.

Measures of the predictive accuracy such as the Accuracy Ratio or Harrell's C are, of course, subject to sampling variability. Analyzing this variation is useful, for confidence intervals and hypothesis tests not least. Methods for statistical inference for the Accuracy Ratio and Harrell's C have been established for approximately independent data

\* Tel.: +49 221 470 6561; fax: +49 221 470 5084.

E-mail address: [walter.orth@uni-koeln.de](mailto:walter.orth@uni-koeln.de).

<sup>1</sup> Some authors focus instead on the ROC curve and its summary statistic, the area under the ROC curve. However, the Accuracy Ratio and the area under the ROC curve contain exactly the same information, since there is a simple linear relationship between the two (Engelmann, Hayden, & Tasche, 2003). The next section describes this relationship explicitly.

(Bamber, 1975; DeLong, DeLong, & Clarke-Pearson, 1988; Newson, 2006a). However, in typical ratings datasets, ratings change over time, which means that we have a time series of default predictions for each obligor. If we want to evaluate the whole time series of default predictions and build cohorts of obligors in time intervals that are shorter than the prediction horizon, we get a sample which consists of partially overlapping data which clearly exhibit dependence. To construct a summarizing index under such a multiple cohort sampling scheme, one may take a weighted mean of the indices of the individual cohorts or calculate the index for the pooled cohort by simply aggregating all of the individual cohorts to form one large cohort (Cantor & Mann, 2003). However, statistical inference for such summarizing indices clearly has to take the dependence of the data into account. We show how this can be done by deriving asymptotic formulae for the weighted mean case, and describe how resampling procedures can be used for both the weighted mean and the pooled version. These procedures use more information than is the case if only a subsample of the data which consists of approximately independent observations is used. In statistical terms, the benefits are narrower confidence intervals and more powerful tests.

In addition to our theoretical considerations, we also provide an empirical illustration of the proposed methods using a dataset of Standard & Poor's long term credit ratings for North American firms. We analyze prediction horizons ranging from 6 to 60 months. One of our main findings is that we observe illusively high values for the Accuracy Ratio at long horizons, as a direct result of the omission of censored observations, i.e. missing data. As will be analyzed in Section 5, the missingness mechanism is such that it leads to an upward bias of the Accuracy Ratios.<sup>2</sup> With respect to the economic implications, this bias may lead to an overestimation of the long-run accuracy of ratings by investors and risk managers who use the Accuracy Ratio. In another part of our empirical study, we provide an example where the aforementioned enhanced testing power changes the test decision.

The methods proposed in this paper are relevant to the development and validation of default prediction models and rating systems. They are useful for all cases in which the prediction horizon covers more than one sample period, and are especially beneficial for the evaluation of multi-period default predictions. On the one hand, multi-period predictions are necessary if longer horizons, say multiple years, are of interest—as is in the case with S&P's long term credit ratings, for instance. The Basel Committee for Banking Supervision has also emphasized the importance of a multi-year perspective, claiming that “banks are expected to use a longer time horizon (than one year) in assigning ratings” (Basel Committee on Banking Supervision, 2006, § 414). On the other hand, a multi-period set-up is also useful for one-year predictions by allowing the use of all of the information in monthly or quarterly data, say. For instance, it is well known that

the analysis of ratings on a yearly basis omits valuable information about intra-year rating transitions (Lando & Skodeberg, 2002).

The remainder of the paper is organized as follows. In the next section, we give a brief review of the Accuracy Ratio and its theoretical background. We then introduce Harrell's C and present an adjusted version of it in Section 3 before we look at statistical inference for the Accuracy Ratio and (adjusted) Harrell's C in Section 4. Section 5 contains the empirical investigation, while Section 6 concludes.

## 2. The Accuracy Ratio

The most popular approach to measuring the predictive accuracy in the rating industry, as well as in the academic world, is based on either the Cumulative Accuracy Profile (CAP) and its summary statistic, the Accuracy Ratio (AR), or the Receiver Operating Characteristic (ROC) curve and its summary index, the area under the ROC curve (AUROC).<sup>3</sup> A comprehensive description of the graphical interpretation of these indices and an overview of further measures is given by Thomas, Edelman, and Crook (2002, Chapter 7). The Accuracy Ratio and AUROC are designed to measure the discriminative power of a rating system. If default probabilities are assigned to ratings, further dimensions of predictive accuracy arise (Krämer & Güttler, 2008). However, we will not pursue this case further here.

We will first focus on the Accuracy Ratio, then provide the link to the AUROC. Besides its graphical derivation, there is another simple method of calculating the Accuracy Ratio which provides some intuition into what this index measures. Denote the ratings (high values indicate a low risk) of the  $i$ th defaulting obligor and the  $j$ th non-defaulting obligor by  $X_i^D$  and  $X_j^{ND}$ , respectively. The numbers of defaulting and non-defaulting obligors in the sample are referred to as  $n_1$  and  $n_2$ , respectively. Define

$$c_{ij} = \begin{cases} 1 & \text{if } X_i^D < X_j^{ND}, \\ -1 & \text{if } X_i^D > X_j^{ND}, \\ 0 & \text{if } X_i^D = X_j^{ND}. \end{cases} \quad (1)$$

Then, the Accuracy Ratio is given by

$$AR = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} c_{ij}. \quad (2)$$

We will call  $c_{ij}$  the concordance score of the pair of the  $i$ th defaulting and the  $j$ th non-defaulting obligor. Concordance is given if the rating of the defaulting obligor was worse than the rating of the non-defaulting obligor, while we have discordance in the opposite case. The case of identical ratings is captured by a concordance score of zero. The concordance score is evaluated for every pair of defaulting and non-defaulting obligors, and then averaged over all possible pairs. It can be shown that the AUROC can also be calculated using Eqs. (1) and (2) by simply replacing

<sup>2</sup> We would like to thank a referee for pointing us to the missing data theory in this context.

<sup>3</sup> The Accuracy Ratio is sometimes referred to as the Gini coefficient.

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات